

Wright State University

CORE Scholar

Computer Science & Engineering Syllabi

College of Engineering & Computer Science

Winter 2011

CS 707-01: Information Retrieval

Krishnaprasad Thirunarayan

Wright State University - Main Campus, t.k.prasad@wright.edu

Follow this and additional works at: https://corescholar.libraries.wright.edu/cecs_syllabi



Part of the [Computer Engineering Commons](#), and the [Computer Sciences Commons](#)

Repository Citation

Thirunarayan, K. (2011). CS 707-01: Information Retrieval. .

https://corescholar.libraries.wright.edu/cecs_syllabi/810

This Syllabus is brought to you for free and open access by the College of Engineering & Computer Science at CORE Scholar. It has been accepted for inclusion in Computer Science & Engineering Syllabi by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.

CS707 Information Retrieval

- **Instructor:** T. K. Prasad
 - **Phone No.:** (937)-775-5109
 - **Email:** t.k.prasad@wright.edu
 - **Home Page:** <http://knoesis.wright.edu/tkprasad>
 - **Quarter:** Winter, 2011.
 - **Class Hrs:** 4:10pm-5:25pm, TTh, Fawcett 218
 - **Office Hrs:** 3-4pm 395 Joshi Research Center (or by appointment)
-

Course Objective

- To cover the foundations of information retrieval, and the design, analysis and implementation of IR systems.
-

Course Prerequisite

- CS600 Data Structures and Algorithms
-

Course Description

This course will cover models for information retrieval, techniques for indexing and searching, and algorithms for classification and clustering. It will also cover SVM, latent semantic indexing, link analysis and ranking, Map-Reduce architecture and Hadoop, to different degrees of detail, time permitting.

Course Load

The course load includes a programming project (30 pts), a midterm exam (30 pts), and a final exam (40 pts).

Required Texts

- "Introduction to Information Retrieval". C.D. Manning, P. Raghavan, and H. Schütze. Cambridge University Press, 2008.

Recommended Texts

- "Modern Information Retrieval". Ricardo Baeza-Yates and Berthier Ribeiro-Neto. Addison Wesley, 1999.
 - "Mining the Web: Discovering Knowledge from Hypertext Data". Soumen Chakrabarti. Morgan Kaufmann, 2003.
 - "Information Retrieval: Algorithms and Heuristics", D. Grossman and O. Frieder. Springer, 2nd Ed., 2004.
 - "Managing Gigabytes: Compressing and Indexing Documents and Images", I. Witten, A. Moffat, and T. Bell. 2nd Ed., Morgan Kaufmann, 1999.
 - "Understanding Search Engines: Mathematical Modeling and Text Retrieval", Michael W. Berry, and Murray Browne, 2nd Ed., SIAM, 2005.
 - "Search Engines: Information Retrieval in Practice", W. B. Croft, D. Metzler, and T. Strohman, Addison Wesley, 2009.
 - "Information Retrieval : Implementing and Evaluating Search Engines", Stefan Büttcher, Charles L. A. Clarke, and Gordon V. Cormack, MIT Press, 2010.
-

Reference URLs

- Lucene Text Search Engine (<http://lucene.apache.org/>)
 - Hadoop (<http://hadoop.apache.org/core/>)
 - Singular Value Decomposition (<http://www.uwlax.edu/faculty/will/svd/index.html>)
 - Strang's Linear Algebra Course (MIT)
(<http://ocw.mit.edu/OcwWeb/Mathematics/18-06Spring-2005/CourseHome/index.htm>)
 - Andrew Moore's Statistical Data Mining Tutorials (CMU)
(<http://www.autonlab.org/tutorials/>)
 - Matei Zaharia's Introduction to MapReduce and Hadoop (Cloud Computing) ([in powerpoint](#)) or ([archived video](#))
-

Grading

The A/B/C/D/F letter grade will be assigned at the end of the course.

Tentative Class Schedule and Syllabus

	Topics	Addl. Reading
Class 1	<u>Information Retrieval; The Boolean Model</u>	MIR-1
Class 2	<u>The Vector Space Model : Term Weighting and Scoring</u>	IIR-6, MIR-2
Class 3	<u>Inverted Index Construction</u>	IIR-1, MIR-8.2
Class 4	<u>Dictionary and Postings; Query Processing</u>	IIR-2, MIR-7.2
Class 5	<u>Tolerant Retrieval (B-Trees)</u>	IIR-3
Class 6	<u>Index Construction</u>	IIR-4, MG-5
Class 7	<u>Map Reduce Architecture</u>	<u>Hadoop</u>
Class 8	<u>Index Compression</u>	IIR-5, MG 3.3-4
Class 9	<u>Vector Space Model : TF-IDF</u>	IIR-6.2-4
Class 10	Midterm Exam (Feb 8)	
Class 11	<u>Vector Space Model : Ranking Revisited</u>	IIR-6.1, IIR-7
Class 12	<u>Evaluation in Information Retrieval</u>	IIR-8, MIR-3 IIR-8
Class 13	<u>Relevance Feedback and Query Expansion</u>	IIR-9, MIR-5.2-4
Class 14	<u>Text Classification and Naive Bayes</u>	IIR-13
Class 15	<u>Vector Space Classification</u>	IIR-14
Class 16	<u>Support Vector Machines</u>	IIR-15, <u>Primer</u>
Class 17	<u>Flat and Hierarchical Clustering</u>	IIR-16, IIR-17
Class 18	<u>Latent Semantic Indexing</u>	IIR-18, <u>Refs</u>
Class 19	<u>Linear Algebra: Matrix Decompositions</u>	<u>SVD-URL</u>
Class 20	Wrap-Up	
Class *	Web Characteristics	IIR-19
Class *	Web Search: Crawling and Indexes	IIR-20
Class *	<u>Link Analysis</u>	IIR-21
	Final Exam (5:45pm-7:45pm, March 15)	

Assignments (Winter 2011)

- Assignment 1. (Sample Medline-1033 dataset)
- Assignment 1 (Alternate)

Exams (Winter 2010)

- Midterm.
- Final.

T. K. Prasad (Jan 3, 2011)